M-DG Seminar: Multinomial Processing Tree Modeling

The Software multiTree

Summer semester 2020

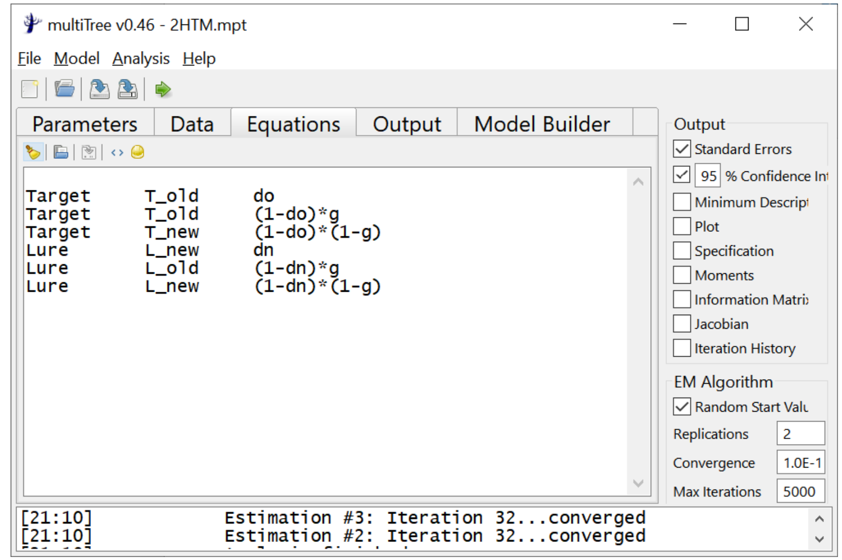
Prof. Dr. Daniel Heck

M-DG: Multinomial Processing Tree Modeling

Part	Date	Topic	Literature
(A) Theory	Self study	A1) Introduction	Erdfelder et al. (2009)
		A2) Basics of MPT modeling	Batchelder & Riefer (1999)
		A3) The software multiTree	Moshagen (2010)
		A4) Hierarchical MPT modeling	Lee (2011) Heck et al. (2018)
(B) Application	15.5.*	B1) Questions & Practice with multiTree	Batchelder & Riefer (1986)
	20.5.*	B2) Workflow: Developing an MPT model	Jung et al. (2019)

* Web-Conference, 12:00 – 15:00, https://webconf.hrz.uni-marburg.de/b/dan-fvk-ha6

Software multiTree



multiTree (Moshagen, 2010)

- Specification, fitting and testing of MPT models
- Freeware
- Requires to install a Java Runtime Environment (JRE)
  - e.g., https://www.java.com/en/download/
- Download:
  - https://www.sowi.uni-mannheim.de/erdfelder/forschung/software/multitree/

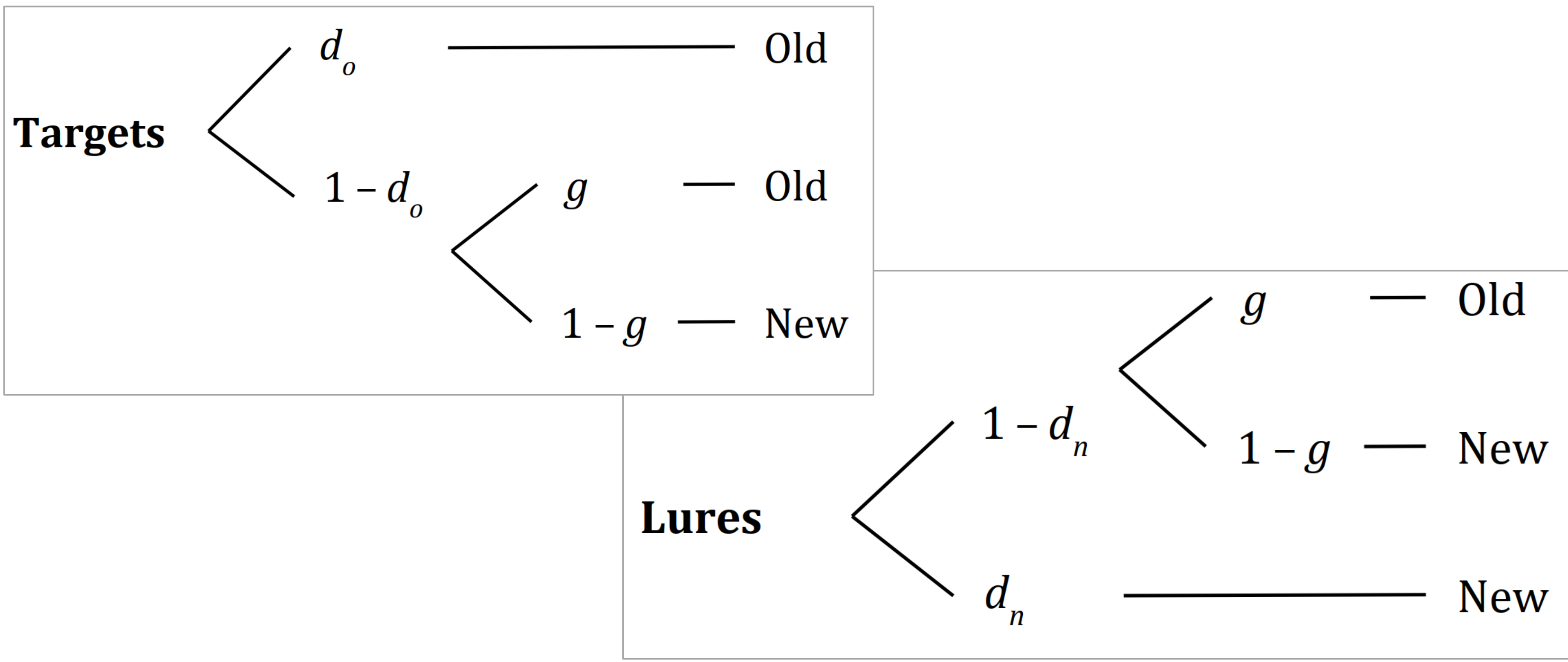
The Software multiTree

Overview:

1. Define MPT model
2. Provide data
3. Check identifiability
4. Comparing Nested Models
5. Preview: Exercises for the first interactive session
6. Advanced: Power analysis

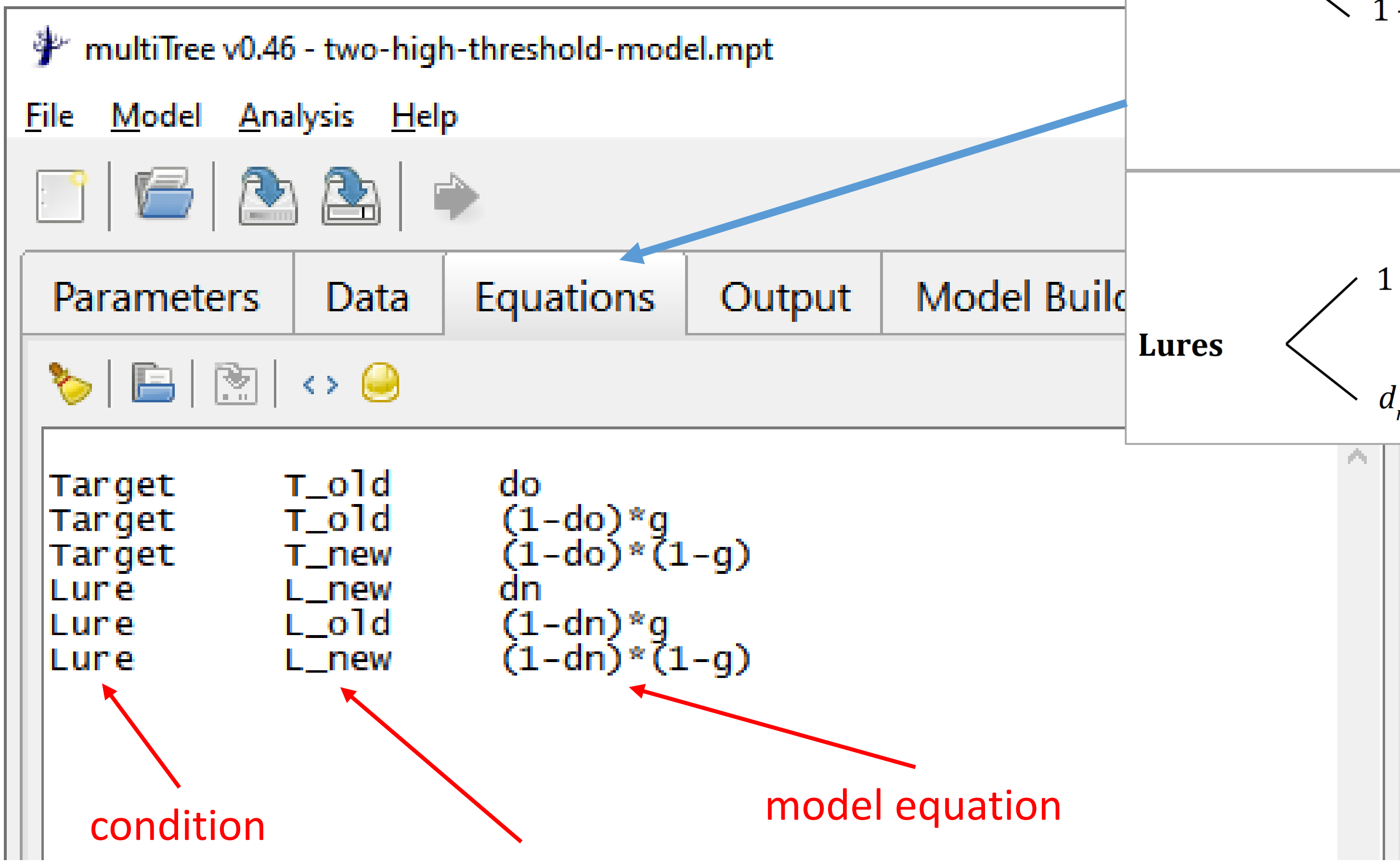
Application: Define MPT Model

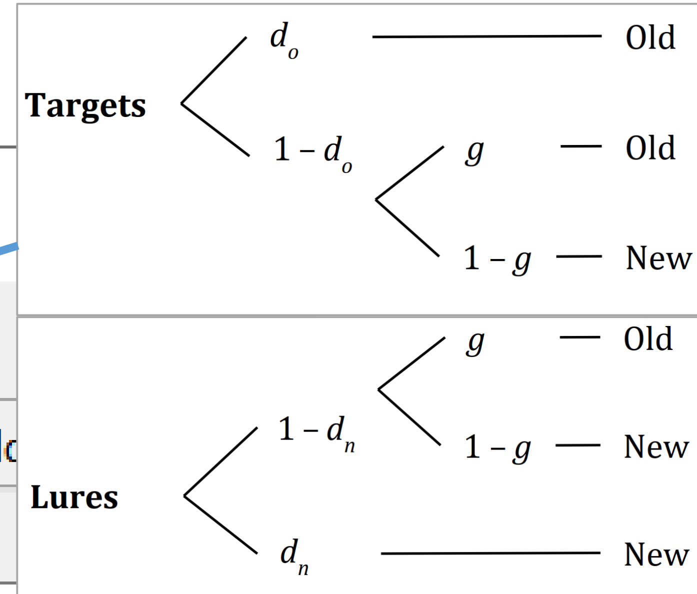
1.) Open multiTree and provide the model equations of the Two-High-Threshold Model:



Two-High-Threshold Model (2HTM)

1.) Specification of an MPT model:





condition

response category

model equation

The Software multiTree

Overview:

1. Define MPT model
2. Provide data
3. Check identifiability
4. Comparing Nested Models
5. Preview: Exercises for the first interactive session
6. Advanced: Power analysis

Application: Provide Data

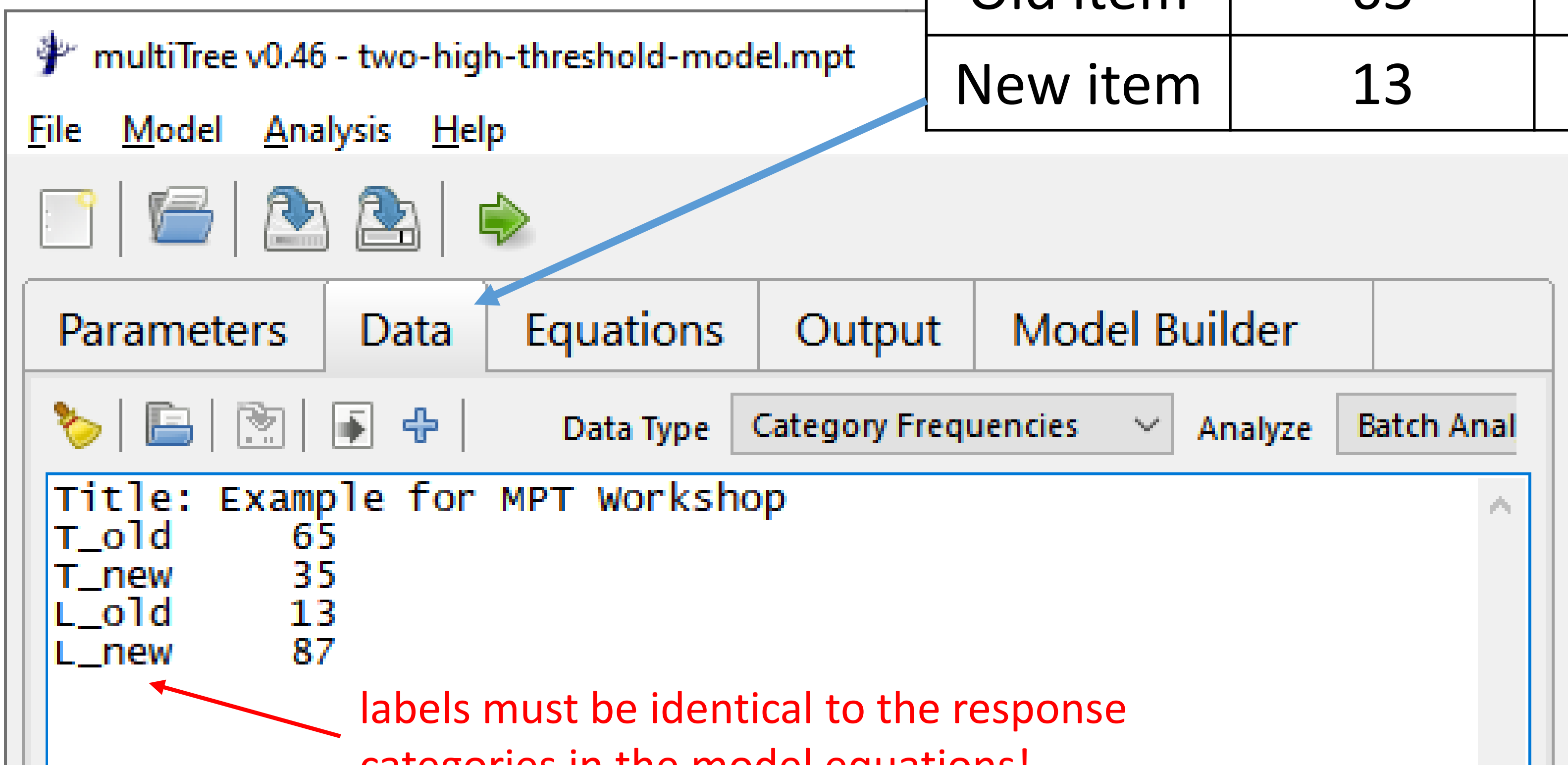
2.) Provide the following data / observed frequencies:

	„old“	„new“
Old item	65	35
New item	13	87

Data

2.) Observed response frequencies:

	„old“	„new“
Old item	65	35
New item	13	87



labels must be identical to the response categories in the model equations!

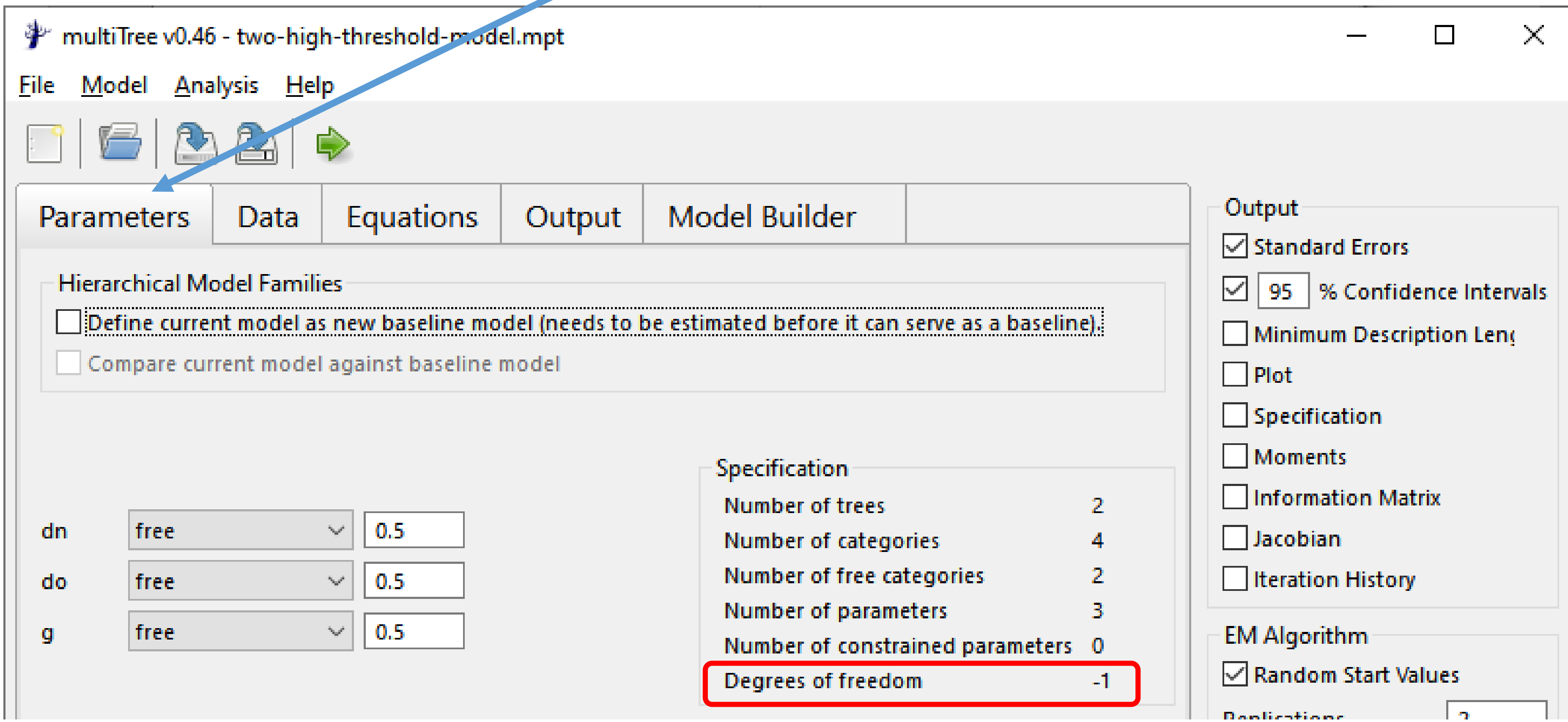
The Software multiTree

Overview:

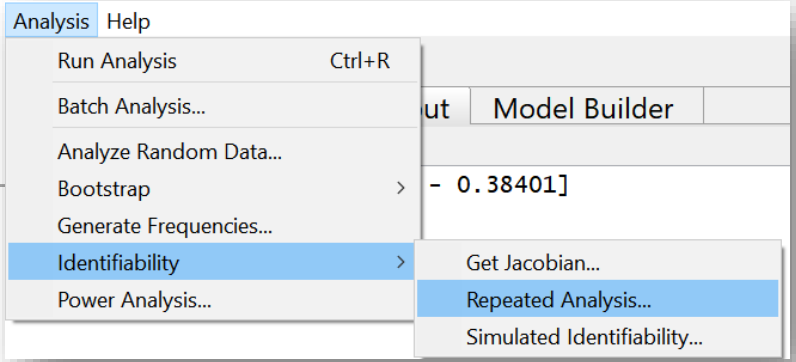
1. Define MPT model
2. Provide data
3. Check identifiability
4. Comparing Nested Models
5. Preview: Exercises for the first interactive session
6. Advanced: Power analysis

Check Identifiability

3.) Is the model identifiable?



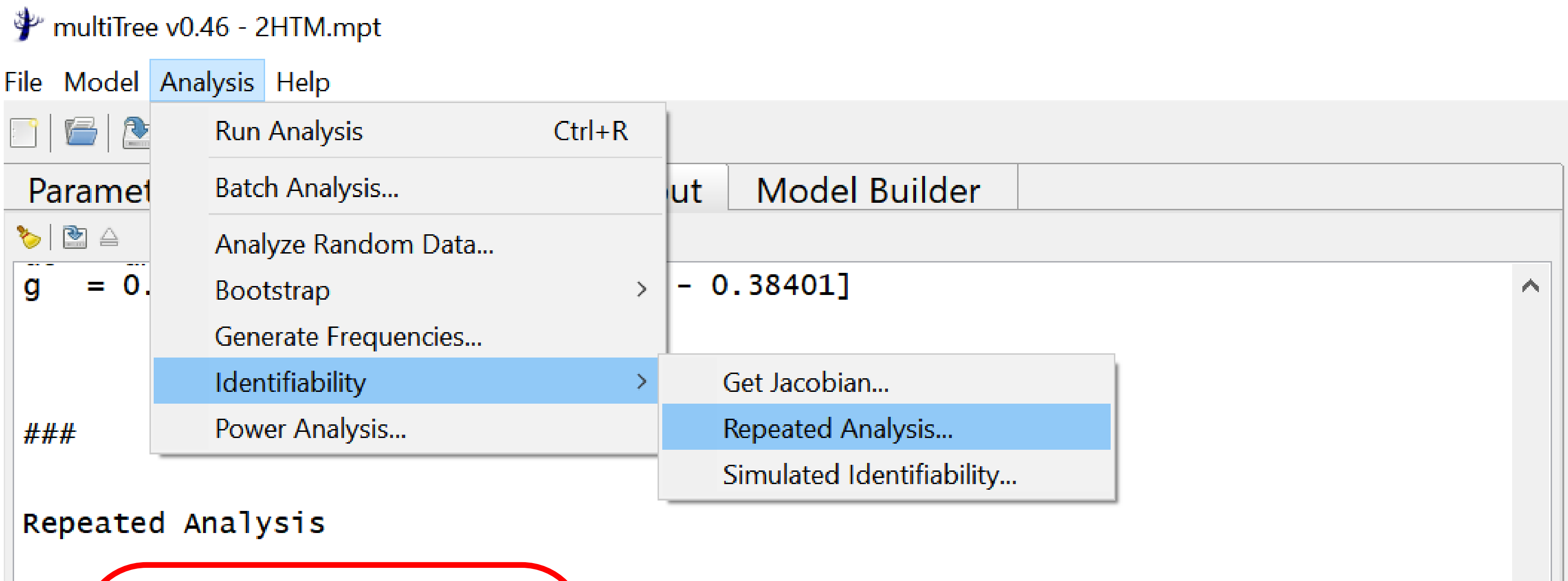
Identifiability Checks in multiTree



- "Get Jacobian"
  - Check whether the rank of the Jacobian matrix (for a random parameter θ) is equal to the number of free parameters
- "Repeated Analysis"
  1. Estimate parameters multiple times for the same data
  2. Check whether parameter estimates are stable/identical
- "Simulated Identifiability"
  1. Draw random parameter vectors θ in Ω
  2. Get expected category frequencies based on MPT model
  3. Estimate the parameters of the MPT model
  4. Check whether estimated & true parameters are identical

Application: Identifiability

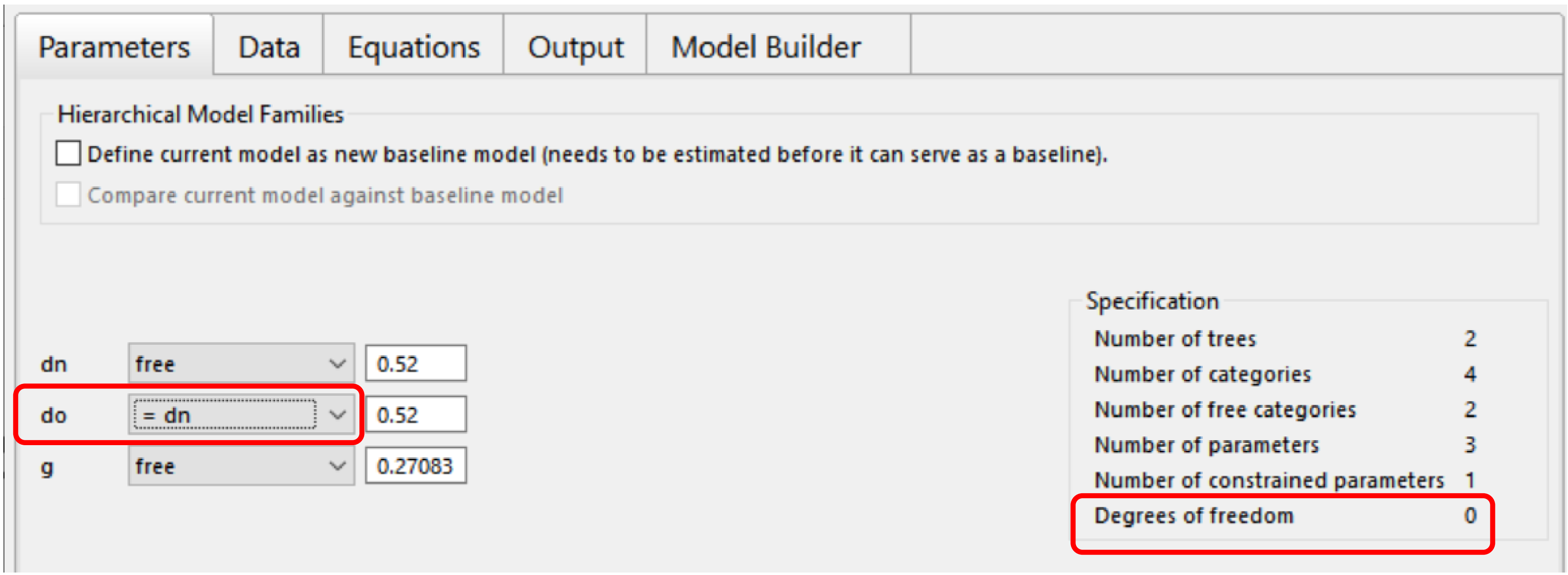
- Example: Check identifiability via repeated analysis



Run	dn	do	g	Fit	AIC	BIC	Delta AIC	Delta BIC
1	0.20114	0.58197	0.16273	0.00000	212.76667	222.66162	2.00000	5.29832
2	0.57178	0.49743	0.30359	0.00000	212.76667	222.66162	2.00000	5.29832
3	0.50071	0.52679	0.26037	0.00000	212.76667	222.66162	2.00000	5.29832
4	0.62998	0.46043	0.35134	0.00000	212.76667	222.66162	2.00000	5.29832
5	0.70466	0.37482	0.44016	0.00000	212.76667	222.66162	2.00000	5.29832
6	0.71662	0.35336	0.45874	0.00000	212.76667	222.66162	2.00000	5.29832
7	0.62593	0.46358	0.34753	0.00000	212.76667	222.66162	2.00000	5.29832
8	0.67560	0.41594	0.40074	0.00000	212.76667	222.66162	2.00000	5.29832
9	0.78259	0.12949	0.59794	0.00000	212.76667	222.66162	2.00000	5.29832
10	0.62316	0.46567	0.34497	0.00000	212.76667	222.66162	2.00000	5.29832

Application: Identifiability

5.) Obtaining an identifiable model by equality constraints:



Model Fitting: Output

6.) Results: Model fit and parameter estimates

Parameters | Data | Equations | Output | Model Builder

```
File:        D:\R\Treebugs_Project\MPT-workshop\2020-02 MPI (Berlin)'
Data Set 1: Title: Example for MPT Workshop


Estimation proceeded normally.


Model Fit


PD^lambda=0.0 (df=0)     = 0.00000

ln(likelihood)           = -103.38334
AIC                      = 210.76667
BIC                      = 217.36330
Delta AIC                = 0.00000
Delta BIC                = 0.00000


Parameter Estimates, Standard Errors, and Confidence Intervals


dn  = 0.52000   (0.05836)   [0.40561 - 0.63439]
do  = dn
g   = 0.27083   (0.05774)   [0.15766 - 0.38401]
```

G² likelihood ratio test

The Software multiTree

Overview:

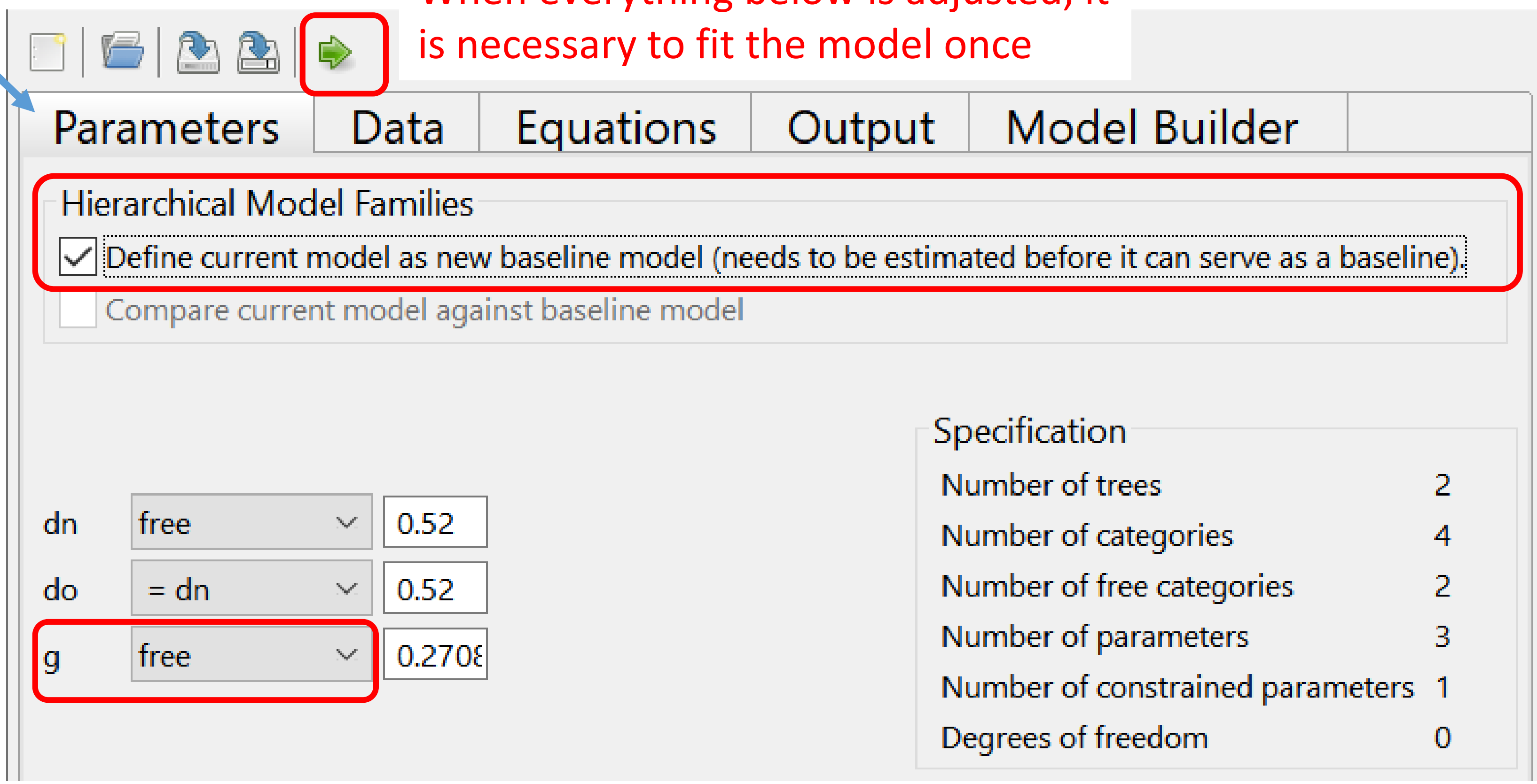
1. Define MPT model
2. Provide data
3. Check identifiability
4. Comparing Nested Models
5. Preview: Exercises for the first interactive session
6. Advanced: Power analysis

Comparing Nested Models

7.) Hypothesis testing: Does the constraint $g = .50$ hold?

→ Step 1: Define baseline model & fit model

When everything below is adjusted, it is necessary to fit the model once

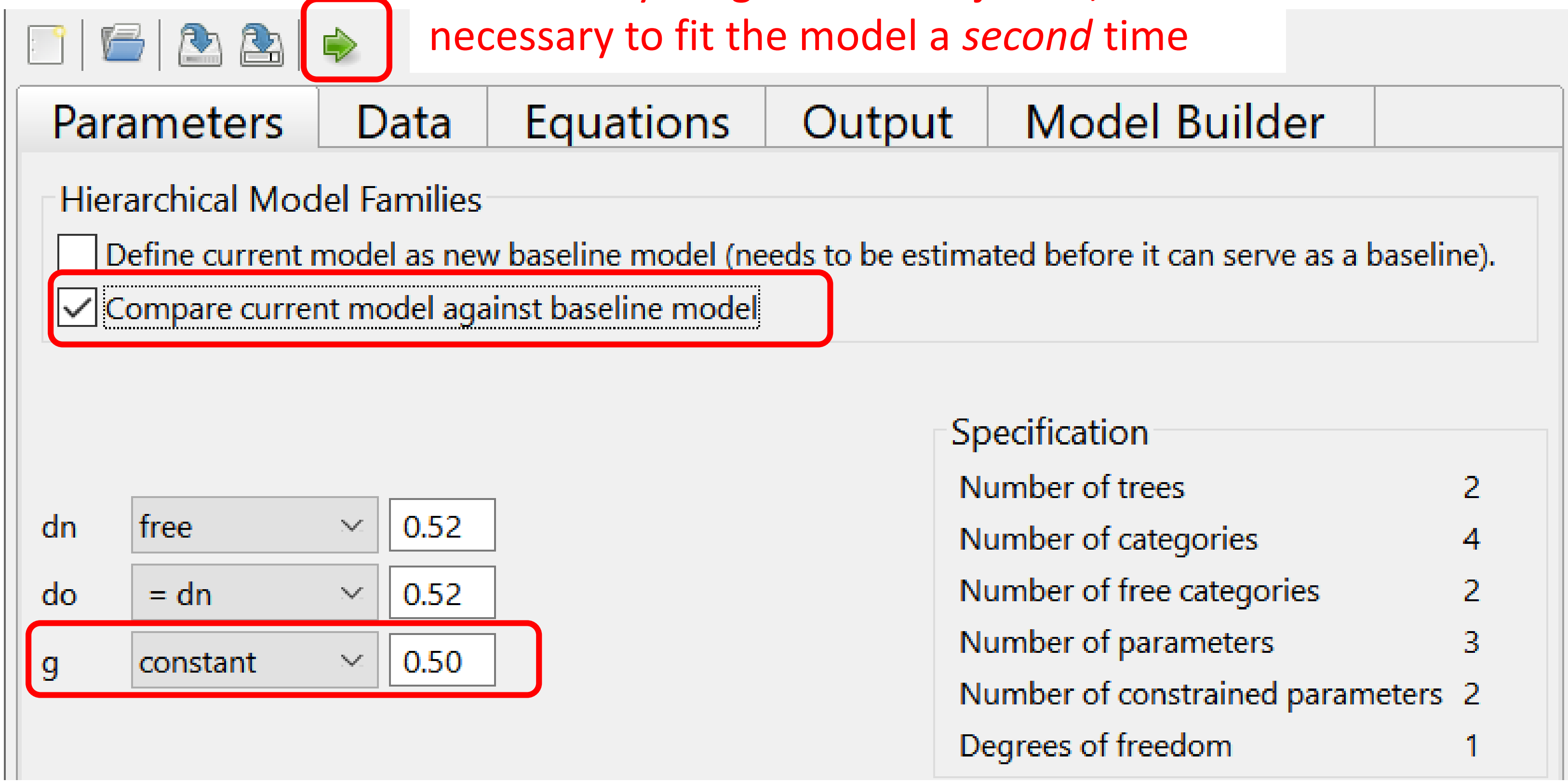


Comparing Nested Models

7.) Hypothesis testing: Does the constraint g = .50 hold?

→ Step 2: Compare current model to baseline

When everything below is adjusted, it is necessary to fit the model a *second* time



Comparing Nested Models

- 7.) Hypothesis testing: Does the constraint $g = .50$ hold?

Likelihood ratio test with ΔG^2 statistic

Parameters | Data | Equations | Output | Model Builder

```
Difference to Baseline Model (Difference = Current - Baseline)

PD^lambda=0.0 (df=1)   = 13.66530      p = 0.00022

AIC difference         = 11.66530
BIC difference         = 8.36699
Ratio of AIC weights   = 0.00292
Ratio of BIC weights   = 0.01502


      Baseline     Current
dn    free         free
do    = dn         = dn
g     free         0.50


Parameter Estimates, Standard Errors, and Confidence Intervals

dn  = 0.52000   (0.06040)   [0.40162 - 0.63838]
do  = dn
g   = 0.50000   (constant)
```

what are the parameters that are tested?

Philipps Universität Marburg

The Software multiTree

Overview:

1. Define MPT model
2. Provide data
3. Check identifiability
4. Comparing Nested Models
5. Preview: Exercises for the first interactive session
6. Advanced: Power analysis

Preview: Exercises for the first interactive session

1. Extend the 2HTM to two base-rate conditions:

		„old“	„new“
30% Targets	Target	65	35
	Lure	13	87
70% Targets	Target	83	17
	Lure	43	57

2. What are the parameters of the extended model?
3. Is the model identifiable with separate d_n and d_o?
4. Fit the model to both conditions jointly.
5. Test whether g differs significantly between conditions.

The Software multiTree

Overview:

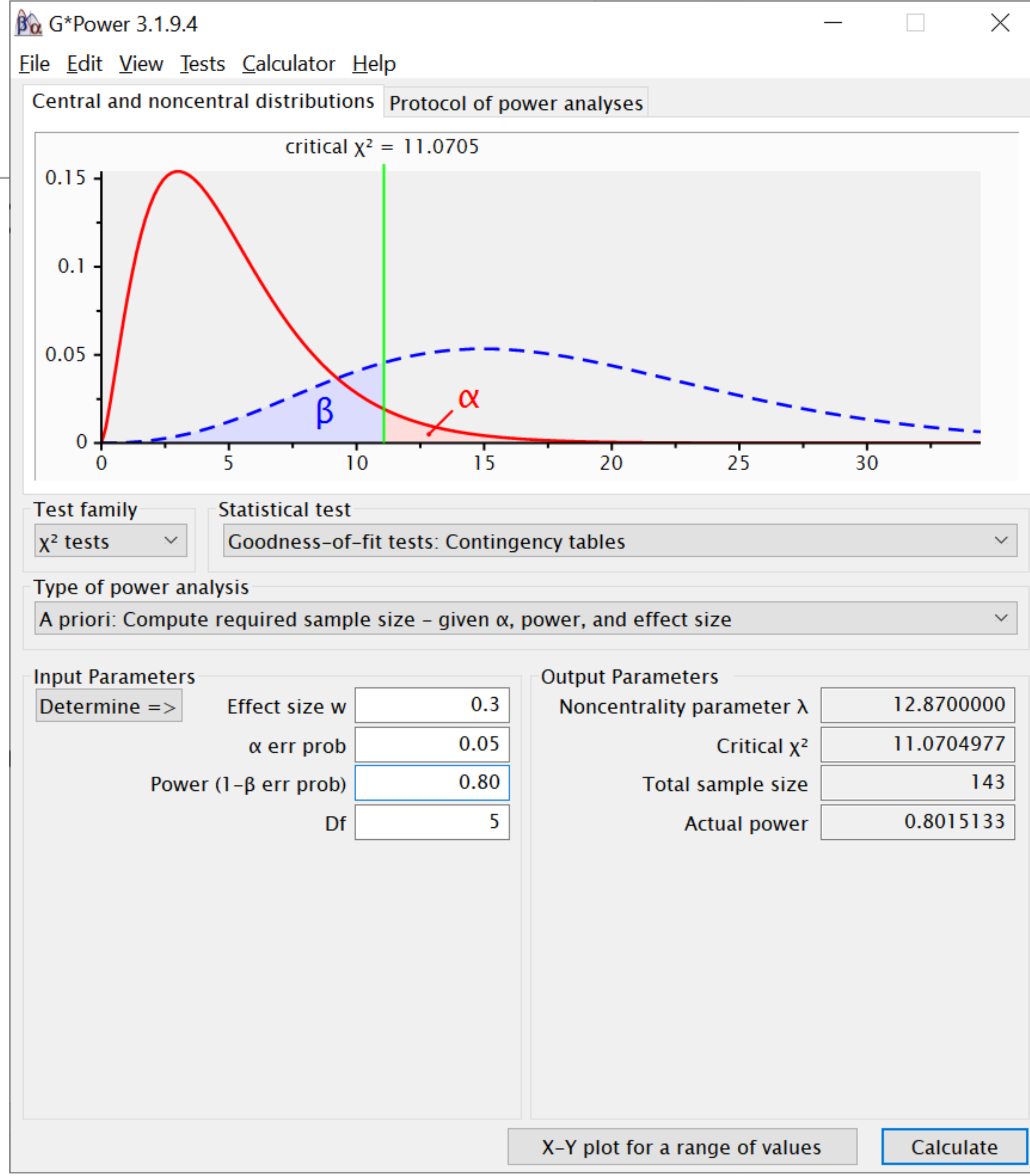
1. Define MPT model
2. Provide data
3. Check identifiability
4. Comparing Nested Models
5. Preview: Exercises for the first interactive session
6. Advanced: Power analysis

Traditional Power Analysis

- Under H_1, the G^2 statistic of H_0 follows a noncentral χ^2 distribution
  - noncentrality parameter $\gamma = N \cdot w^2$
  - w denotes the effect size ($w^2 = G^2(H_0 \mid H_1$ holds given $\boldsymbol{\theta}) / N$)
- Effect size conventions
  - $w = .10$ ("small effect")
  - $w = .30$ ("medium effect")
  - $w = .50$ ("large effect")
- Types of power analysis
  - A priori: Compute required N as a function of w, α, and $1 - \beta$
  - Post hoc: Compute $1 - \beta$ as a function of w, α, and N

Traditional Power Analysis

- Example with G*Power for a “medium” effect size (w = 0.3)



Traditional Power Analysis: Limitations

- Problems:
  1. How does effect size translate into parameter values?
  2. Same meaning of effect size labels in different models?
  3. Relative size of groups/conditions is ignored.

- Cohen (1988, p. 244) on *w* effect sizes conventions:
  - "Their use requires particular caution, since, apart from their possible inaptness in a particular substantive context, what is subjectively the same degree of departure or degree of correlation (…) may yield varying *w*, and conversely. The investigator is best advised to use the conventional definitions as a general frame of reference (…) and not to take them too literally."

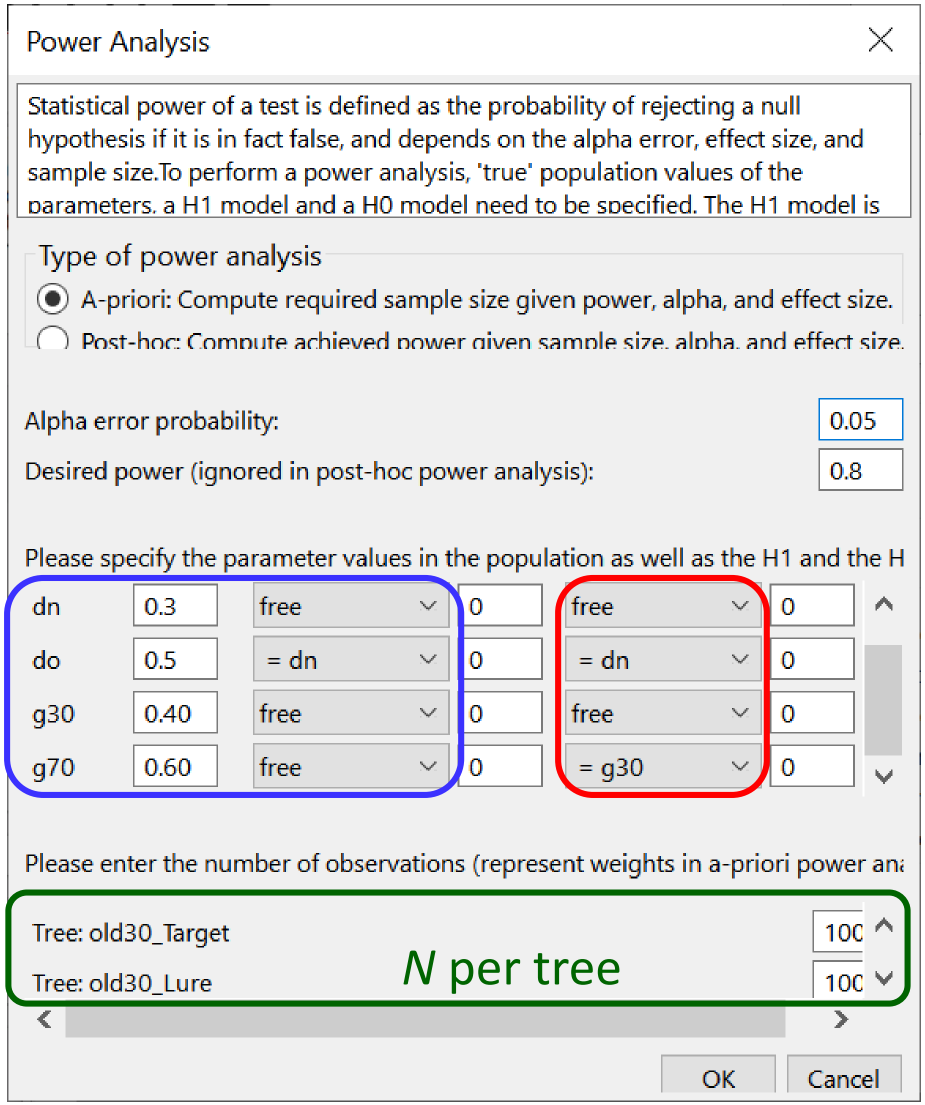
Approach 2: Meaningful Power Analysis

Power as a function of the model parameters $\boldsymbol{\theta}$ under H_1:

1) Specify a specific H_1 model with all parameter values $\boldsymbol{\theta}$ fixed at „plausible values“
2) Specify a nested H_0 model
3) Choose number of observations N_k for each tree *k* and calculate the expected frequencies under H_1
4) Fit the H_0 model to the H_1 expected frequencies by minimizing G^2
5) Use minimum G^2 value as noncentrality parameter γ
6) Compute the power $1 - \beta = P(\chi^2(\gamma, df) \geq c_{(df,\ \alpha)})$

Power Analysis in multiTree

H_1: "true state of the world"
→ power may depend on the values of *all* parameters **θ**!



H_0: nested model
→ which parameter constraints are tested
→ here: $g30 = g70$

Power Analysis in multiTree: Output

Parameters | Data | Equations | Output

```
Model specifications

             Population     H1       H0
dn         = 0.30000        free     free
do         = 0.50000        = dn     = dn
g30        = 0.40000        free     free
g70        = 0.60000        free     = g30


Required number of observations = 460.72733
Required observations per tree:
             Tree: old30_Target =  115.18183
               Tree: old30_Lure =    115.18183
            Tree: old700_Target = 115.18183
              Tree: old700_Lure =   115.18183


Model Fit
H0 Model PD^lambda=0.0 (df=2)   = 8.07460
H1 Model PD^lambda=0.0 (df=1)   = 0.22574


Effect size w                   = 0.13052
Non-centrality parameter        = 7.84886
Critical Chi-square             = 3.84146
Delta df                        = 1


Alpha error probability         = 0.05000
Requested power                 = 0.80000
Achieved power                  = 0.80000
```

Result:
→ How many observations are required?